|  |  |
| --- | --- |
| Slide | Description of Content |
| 1 | Identify clear learning targets   |  | | --- | | Generalizations and principles | | Key concepts and core processes | | | | Facts and skills | | | | | |
| 2 | What is a discipline?  “Many people erroneously thing of academic disciplines as the content, but that is NOT what a discipline is. Science is a discipline because the habit of jumping to conclusions based on prior beliefs runs deep in human beings (and novice scientists), and is overcome only through the discipline of trying to isolate key variables, and methodically testing for them. You have to learn the discipline of carefully observing, gathering apt evidence, and weighing its limited implications while remaining skeptical. This so-called scientific method is not a skill, but a set of dispositions, skills, and transfer abilities in the use of content, learnable only by doing. Similarly, the goal in learning to do history is to avoid present-centeredness and simplistic casual reasoning. One must learn to think and act like a journalist/ curator/ historian to learn the discipline of history. Learning only the factual content or highly-scripted skills is as little likely to make you disciplined as merely practicing discrete moves in basketball will equip you to be a successful game player.” Wiggins and McTighe, *Schooling by Design* (2007) |
| 3 | Feedback from assessments should provide students with a clear picture of their progress on learning goals and how they might improve:  Bangert-Drowns, Kulik, Kulik, & Morgan, 1991   |  |  |  | | --- | --- | --- | | # of studies | Characteristic of Feedback from Classroom Assessment | Percentile Loss or Gain | | 6 | Right/wrong | -3 | | 39 | Provide correct answers | 8.5 | | 30 | Criteria understood by student vs. not understood | 16 | | 9 | Explanation | 20 | | 4 | Student reassessed until correct | 20 | |
| 4 | Fuchs & Fuchs, 1988: Evaluation by rule (uniform way of interpreting results of classroom assessments using a tight logic) 49 studies, 39 percentile point gain;  The “grade” tells a student exactly what they know or don’t know vs. “a 72% means a C” |
| 5 | Three types of assessment: obtrusive, unobtrusive, student-generated  Three uses of assessment: formative scores; summative scores; instructional feedback |
| 6 | Formative scores are recorded and tracked but NOT used to determine a grade;  Summative scores represent a student’s final status after some interval of time (not necessarily a “test” at the end);  Instructional Feedback is not recorded and is used to provide students and teachers with information that should change their behavior |
| 7 | What is wrong with the 100 point scale?   |  |  | | --- | --- | | Items 1-10  Require recall of important but simpler content that was explicitly taught | Total for Section: | | Items 11-14  Ask for application of complex content that was explicitly taught AND in situations similar to what was taught | Total for Section: | | Items 15-16  Ask for application in new situations that go beyond what was explicitly taught | Total for Section | |
| Activity: Have participants decide how much weight to give each of the three categories in the previous row, so that the total adds up to 100.  Next, imagine a student scored 8/10 in the first section, 2/4 in the second section, and 0/2 in the third section. Have participants add up the total points they would give that student based on their weighting.  Describe the 4 point scale below and then revisit the scoring on the assessment above. Now Items 1-10 represent a “2”, Items 11-14 represent a “3” and Items 15-16 represent a “4”. How does the score change now?? | |
| 8 | Instead use 4 point scale with explicit performance descriptors:   |  |  | | --- | --- | | 4 | In addition to exhibiting level 3 performance, in-depth inferences and applications that go BEYOND what was taught in class (ADVANCED LEARNING GOAL) | | 3 | The student’s responses demonstrate no major errors or omissions regarding any of the information and/or processes that were explicitly taught (TARGET LEARNING GOAL) | | 2 | No major errors and omissions regarding the simpler details and processes, but major errors and omissions regarding the more complex ideas and processes (SIMPLER LEARNING GOAL) | | 1 | With HELP a partial knowledge of some of the simpler and complex details and processes (PARTIAL CREDIT WITH HELP) | | 0 | Even with help, no understanding or skill demonstrated (EVEN WITH HELP, NO SUCCESS) |   The score point “3” describes the target learning goal. Vocabulary and recall is typically a level “2”.  Students who don’t complete the assignment are not scored – this is NOT a “0”. |
| 9 | On an assessment, there are three types of items:  Level 2: simpler details and processes that have been explicitly taught  Level 3: complex ideas and processes that have been explicitly taught  Level 4: inferences and applications that go beyond what was taught  Items are correct, incorrect, or partially correct (low or high) |
| 10 | Have students flesh out the meaning of the scale so they understand exactly what is expected of them  Students co-construct criteria:   * Learning becomes more explicit * Learners can confirm, consolidate, and integrate new knowledge * Scaffolds future learning * Teaches learners what quality looks like * Helps learners learn the language of assessment |
| Discuss the idea of partial credit to earn “0.5” points on the 4 point scale. Partial low, partial high, etc. | |
| 11 | Conversion to percent:   |  |  | | --- | --- | | 4.0 | 100% | | 3.5 | 95% | | 3.0 | 90% | | 2.5 | 80% | | 2.0 | 70% | | 1.5 | 65% | | 1.0 | 60% | | Below 1.0 | 50% | |
| 12 | The scales could be developed for each student-centered learning outcome that has a common assessment or series of assessments. “Unpacking” the learning outcomes will provide the criteria. |
| 13 | Purposes of Assessment:   * What are the major reasons we use report cards and assign grades to students’ work? * Ideally, what purposes should report cards or grades serve? * What elements should teachers use in determining students’ grades? |
| Have participants answer the questions above regarding the purposes of assessments (individually). Discuss as large group. | |
| 14 | Purposes of Grading   * Communicate the achievement status of students to their parents and others * Provide information for student self-evaluation * Select, identify, or group students for certain educational programs * Provide incentives for students to learn * Document students’ performance to evaluate the effectiveness of instructional programs * Provide evidence of students’ lack of effort or inappropriate responsibility   You can’t serve all these purposes with just one instrument. Define the purpose first and let the purpose drive your methods. |
| 15 | Conclusions from the research:  Grading and reporting are NOT essential to the instructional process. Checking is essential – teacher is an advocate; regular checks on student progress paired with guidance and feedback for improvement; what has been learned well, what hasn’t, and what can you do about it? |
| 16 | Conclusions from the research:  No one method of grading and reporting serves all purposes well; method of grading follows purpose   * 4 Methods: Letter grades, percentage grades, standards-based, narrative * Methods can be combined to enhance their communicative value (i.e. standards-based and letter grade/percentage example by Marzano above) * First grading period sets the stage for the entire year – use this to show students they can succeed   Solution   * Determine the primary purpose of each grading and reporting tool * Select or develop the most appropriate method for each tool * Develop a multifaceted, comprehensive reporting system |
| 17 | Conclusions from the research:  Grading and reporting will always involve some degree of subjectivity; detailed and analytic reports are better learning tools |
| 18 | Conclusions from the research:  Mathematic precision does NOT yield fairer or more objective grading;  Questionable practices:   * Averaging to obtain a course grade: hurts kids motivation because one low grade (or ZERO) hurts you for the rest of the term * Giving zeros for work missed or turned in late: use “incompletes” as part of RTI instead with immediate, systemic support * Taking credit away from students for infractions: i.e. homework can add to your grade not take away from (Maureen’s example) |
| Give some sample integer scores for sample students:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Student 1 | Student 2 | Student 3 | Student 4 | Student 5 | | 1 | 0 | 3 | 2 | 3 | | 2 | 1 | 3 | 3 | 2 | | 3 | 2 | 3 | 4 | 2 | | 3 | 3 | 3 | 3 | 4 |   Have people discuss what “grade” they would give each student  Discuss averaging and giving zeros then have participants come back and discuss these five students’ scores again | |
| 19 | Alternatives to Averaging   * Give priority to most recent evidence: i.e. “if you do better on the comprehensive final than on the midterm, we’ll drop the midterm” * Give priority to most comprehensive evidence: i.e. “if you do better on the comprehensive unit assessment, then we’ll drop the quizzes” * Give priority to evidence related to the most important learning goals or standards (use Marzano’s system to develop assessments that are standards-based, aligned to the learning goals) |
| 20 | Alternatives to Giving Zeros   * Assign an incomplete or in progress grade with specific and immediate consequences and support * Report behavioral aspects separately (i.e. separate the product and the process) * Change grading scale to use integers (see Marzano’s example); in 100 point system an “F” is 0-60; in integer system an “F” is 0-1 |
| 21 | Conclusions from research:  Grades have some value as rewards, but NO value as punishments |
| 22 | Conclusions from research:  Grading and reporting should always be done in reference to learning criteria, never on the “curve”   * Product criteria: doesn’t worry about how they got there, just concerned with where they end (i.e. what can they do with respect to the criteria) * Process criteria: how did students get there? (i.e. participation, homework, etc) * Progress criteria: growth model   No evidence to support that one is better than another; separate reporting criteria for each category is best |
| 23 | Guidelines for better practice:   1. Begin with a clear statement of purpose    1. Why are grading and reporting done?    2. For whom is the information intended?    3. What are the desired results? 2. Provide accurate and understandable descriptions of student learning.    1. More a challenge in effective communication    2. Less an exercise in quantifying achievement 3. Use grading and reporting to enhance teaching and learning    1. Facilitate communication between teachers, parents, and students    2. Ensure efforts to help students are harmonious |
| 24 | Quality Classroom-Based Assessment |
| 25 | Typical Assessment Plan   * National Assessments * State Assessments * District Assessments * Common Team Assessments * Classroom Assessments   Questions to Ask:   1. Why do we give these? 2. What data do we get? 3. Are they value added? Cost/benefit?   DMPS Has/DMPS Needs |
| 26 | Balanced Assessment System   |  |  |  | | --- | --- | --- | | Large Scale  Assessment of… | Mid-Scale  Assessment for… | Small-Scale  Assessment for… | | * Summative in nature * Norm-referenced * Aptitude * Achievement | * Formative processes * Criterion-referenced * Often teacher or district-made * Achievement | * Questioning * Day by Day, minute by minute * Achievement | | Essential Question:  What have students already learned as compared with others? | Essential Question:  How can we help students learn more? | Essential Question:  How can we help students learn more? | |
| 27 | Why is reliability difficult?  250 Studies:   * Classroom-based assessments have a reliability of .45 * National and statewide assessments have a reliability of .75-.85   Why the difference?  Teachers weigh/rank assessment items differently over same content; too hard/too easy questions; etc |
| 28 | Six Quality Criteria for Writing Assessments   1. Assessments reflect the essential learnings and standards 2. The students have had the opportunity to learn the content 3. Assessments are as free from bias as possible 4. Assessment levels are appropriate 5. There is consistency in scoring 6. The mastery levels are appropriate |
| 29 | Assessments reflect the essential learnings and standards: Proficiency level descriptions clearly delineate what knowledge and skills students demonstrate for varying levels of understanding (i.e. beginning, progressing, proficient, advanced)  By categorizing students according to CLEAR proficiency descriptions, you can determine appropriate interventions for kids who are not yet proficient (i.e. if students are “beginning” they need different interventions than if they are “progressing”  Create the scale BEFORE you look at assessment items  Put easier items at the front to give kids some confidence |
| 30 | Sufficiency –   * Has depth (enough items to show learning has taken place and students are not just guessing) * Has breadth (range of coverage or difficulty levels to show learning) * Includes all four proficiency levels for coverage   Low stakes: multiple assessments are used to make decisions about groups of students (3 items per proficiency level = 12 total)  High Stakes: one assessment is used to make decisions about INDIVIDUAL students (i.e. exit exams, placement exams, etc) (6-8 items per proficiency level = 24-32 total) |
| 31 | Validity   * Assessments reflect the essential learnings and standards   Fairness   * Students have an opportunity to learn * Assessments are as free from bias as possible * Assessment level is appropriate for students   Test fairness (bias) by asking kids to write on back of the test if there was any item, word, etc that they didn’t understand/hadn’t seen before |
| 32 | Reliability and Validity   * Neither reliable or valid – students’ scores are all over the place * Reliable but not valid – students’ scores are in the same area, but not near the target * Reliable and valid – students’ scores are consistent and near the target (i.e. a “bulls-eye” in darts) |
| 33 | Strategies for Increasing Reliability and Validity   1. Retest: studies show there is a negligible amount of “residual memory” when kids take the exact same test again (may be appropriate to pre-post test kids after all…) 2. Split half: items at various scale levels are split among two tests given at different times (2 beginner items on one test, 2 similar content beginner items on another, etc); if student gets one right on one test, but not another, then it’s a ‘0’, if they get both right, it’s a ‘1’; go item by item, add up the number of 1s, divide by total to get the % reliability |
| 34 | Data Analysis  If we believe that all students can achieve, then any subgroup we choose should have similar achievements and results. |
| 35 | Process for Data Analysis   1. Set the structure 2. Collect the data 3. Analyze Patterns\* 4. Factual Observations – NO inferences 5. Pose hypotheses 6. Accept or refute hypotheses 7. Develop or refine goals for upcoming school improvement cycle 8. Design or review strategies and develop a shared understanding about quality content and specific practices 9. Define and refine evaluation |
| 36 | Data Collection Categories   * Student achievement: standards data, state assessment, others? * Demographic data: enrollment, sub-groups, characteristics of students, etc * Program patterns: student-teacher ratio, course enrollment, special programs, staff development, etc * Perception patterns: climate; parents, student, and staff surveys; safety; attendance; etc   All four types of data are needed to tell the whole story, not just one! (Baseball analogy – score does not tell the whole story of the game) |
| Here might be a good place to do the activity with the 4 types of data:  List sources of data in each of the four categories on separate, color-coded strips of paper  Have groups identify whether or not they have access to each piece of data  Illustrates how we have a lot of data, but not much information (DRIP) | |
| 37 | Levels of Data Analysis   1. Snapshots 2. Snapshots over time 3. Two or more variables within the same area (i.e. one of four categories of data) 4. Two or more variables within the same area over time 5. Intersection of two categories of measures (This is where you can begin to examine casual relationships) 6. Intersection of two categories of measures over time 7. Intersection of three categories of measures 8. Intersection of three categories of measures over time 9. Intersection of four categories of measures 10. Intersection of four categories of measures over time   *Data Analysis for Continuous School Improvement (2nd ed.)* |
| 38 | Organizing factual observations, hypotheses, and questions: Make a chart   * List strengths in the format of factual statements * Note any questions that arise * List observed challenges in the form of factual statements * Chart further questions and hypotheses (theory or possible explanation for the patterns that are observed) * Seek further data to accept or refute the hypotheses  |  |  | | --- | --- | | Strengths: | Challenges: | | Hypotheses:  Evidence to accept or reject hypotheses | Questions |   Critical friends review   * How easy are data charts to read and understand? * What questions would you pose from a glance at the data? * What works, what could be improved? |
| 39 | Criteria for Evaluation of a Student Performance Goal   * Goals are based on analysis of data * They are supportive of the district and school vision * They are phrased in terms of student learning * They address all of the school’s learners * They reflect an emphasis on student growth * Goals can be implemented school wide * There is faculty consensus on the essence or definition of meaning for each goal * **Goals are supported by a minimum of three data sources** |